Integrating animal movement with habitat suitability for estimating dynamic landscape connectivity

Mariëlle L. van Toor, Bart Kranstauber, Scott H. Newman, Diann J. Prosser, John Y. Takekawa, Georgios Technitis, Robert Weibel, Martin Wikelski, Kamran Safi

Electronic Supplementary Materials

A Segmentation by habitat use and suitability landscapes

The requirements of individual bar-headed geese are not constant across the year or for different life-history stages. Similarly, the environments they inhabit fluctuate across these periods, and so does the availability of resources. Consequently, the ecological likelihood of a specific trajectory should also be subject to change. To be able to account for these changes, we used a segmentation by habitat use to define periods of the year during which both are consistent (van Toor, Newman, *et al.*, 2016). In conjunction with time series of environmental information, we used these periods to derive several models of habitat use and their respective predictions which reflect the temporal variation in habitat suitability. We mainly followed the procedure that was described in detail by van Toor, Newman, *et al.*, (2016), and we here give an outline of how we derived temporally explicit predictions of habitat suitability. The segmentation by habitat use is a procedure that bins environmentally annotated animal location data into arbitrarily small windows of window size ws, and uses these bins to model habitat use as a series of random forest models (Breiman, 2001). Subsequently, the bins are clustered according to the similarity of habitat use represented by the respective models respresenting segments of consistent habitat use. These segments can then be used to build models that correspond to different habitat use and times of the year. As an extension of the concept of species distribution modelling, it requires presence as well as pseudo-absence data of the species to be modelled, and environmental predictors of habitat suitability. As this approach, different from more classic species distribution modelling approaches, identifies changes in habitat use in time, the environmental predictors are required to be explicit in space as well as in time.

We used the tracking data of all individuals as presence data for the segmentation by habitat use, including the ARGOS Doppler-shift locations that were excluded for the simulation of trajectories. We accounted for the differences in spatial error of GPS and Doppler-shift locations using

dynamic Brownian Bridge Movement Models (Kranstauber, Kays, *et al.*, 2012) as described by van Toor, Newman, *et al.*, (2016), and sampled pseudo-presences from the resulting utilisation distributions. As true absence data were not available to use, we sampled pseudo-absence data from the utilisation distributions of randomised trajectories (see van Toor, Newman, *et al.*, 2016 for details). We used the Movebank Env-DATA system (Dodge, Bohrer, *et al.*, 2013) to annotate the resulting location dataset with environmental information (see Table S 1 for the complete list of environmental variables that were used for the initial models), and used the available interpolation methods to reflect the conditions at the respective locations and time. In addition to the environmental products available at the Env-DATA system, we complemented the data with few additional variables.

To prepare the data for segmentation, we subsetted the tracking data for each species using Julian days (i.e., irrespective of the year). We used four different initial window sizes (5 days, 10 days, half month, month), and built random forest models for all subsets that contained at least 100 locations. For each of the replicates, we randomly split the subset data into 50% test and training data sets, and built a first model with all environmental predictors on a random sample of maximum 2000 training locations. We then used a nested cross validation by sequentially reducing the number of predictors, and repeated this process five times to determine the best number of predictors (n_{opt}). For the final random forest model, we determined the importance of each predictor, and only used the n_{opt} most important predictors. We calculated transferability matrices for each subset and replicate as well as a time distance matrix for the subset Julian dates according to van Toor, Newman, *et al.*, (2016) and used a constrained correspondence analysis to ordinate the matrices. We determined the cluster silhouettes (Rousseeuw, 1987) after fuzzy clustering with the number of clusters $k = 1, ..., 12$, and used the k that produced the most compliant hard clustering. Subsets were then annotated with the respective segment id according to the clustering. We applied a ranking approach to determine the best initial window size and segmentation. We ranked each of the four window sizes according to five criteria: a) the number of segments (more is better), b) the number of switches between segments corrected for the number of segments (less is better), c) the number of unique Julian days covered by the segments (more is better), d) the size (with respect to days) of the initial size (less is better), and e) whether the first segment of the year corresponded to the last segment. Ties were given an average rank. We chose the segmentation with the highest average ranking, and in the case that two window sizes achieved the same rank, we chose in favour of the segmentation that had achieved a higher rank with respect to to criterion c). We then repeated

the modelling procedure for the segmented data with 100 replicates per segment. We calculated model performance of each of the model replicates on the test data using point-biserial correlation (Liu, Berry, *et al.*, 2005), and determined the upper 10%-quantiles to keep only the best ten model replicates per segment.

For the present study, the initial window size of half-months, or 15 days, was given the highest rank. This segmentation had resulted in five distinct clusters of habitat use (see Figure S 3), which we subsequently used to derive suitability landscapes. We prepared daily layers containing the environmental conditions during that day using the same variables as used for the segment models, layed out on a spatial grid with cells of $0.25x0.25$ degrees. We assigned each of these layers with the corresponding model replicates and subsequently derived predictions from each model replicate. First, we averaged over the predictions made from the model replicates, and then calculated the mean ± s.d. for each grid cell for all layers in each segment. We standardised the averages of the resulting five maps, and used these as suitability landscapes for calculating the benefit of staying at stopover sites, b_k.

Table S 1. A list of the environmental variables that were used for the random forest models to predict habitat suitability. While all variables were included in the initial models, a variable selection process reduced the number of variables present in the final models.

Product	**Spatial resolution**	**Source**
Aster GDEM (elevation model)	1 arc-second	`asterweb.jpl.nasa.gov`
Global climate teleconnection	none	`cpc.ncep.noaa.gov`
eTopo1 (ice surface)	1 arc-minute	`ngdc.noaa.gov`
GlobCover 2009 (land cover)	30 arc-seconds	`dup.esrin.esa.it`
MOD13A1.005 (NDVI)	500 meters	`lpdaac.usgs.gov`
MOD11A2.005 (surface temperature)	1000 meters	`lpdaac.usgs.gov`
MOD10A2.005 (Snow products)	1000 meters	`lpdaac.usgs.gov`
Distance to nearest coast	0.01 degrees	`oceancolor.gsfc.nasa.gov`
Distance to nearest lake		derived from Wessel and Smith, (1996)
Ecoregions		Olson, Dinerstein, *et al.*, 2001
Global lakes and wetlands database		Lehner and Döll, 2004

B Estimated flight times

Our measure of ecological likelihood requires an estimate for the duration of migratory trajectory j if individuals didn't use stopovers when migrating from a and b, $\tau_{a,b}$. While τ for a simulated trajectory can be approximated by excluding those parts that are in stopover mode, the first and last position of a stopover will likely not be the same. As this will introduce a minor bias in the expected flight time τ, we decided to rather derive a model from the empirical data. Here, we used that migratory bouts that we had identified previously to derive the measures necessary for the transition probabilities. Each migratory bout presented a series of consecutive locations within the trajectory of an individual bar-headed goose that were classified as fast movement by the EMbC (Garriga, Palmer, *et al.*, 2016). We calculated the geographic distance as well as the time difference between the first and last location of all migratory bouts. We used these time differences as the dependent, and the geographic distance as the independent predictor in a simple linear model. This resulted in a function of $\tau_{a,b} = 0.07471 \times d_{a,b}$, where $d_{a,b}$ is the geographic distance between a and b meters, and $\tau_{a,b}$ the predicted flight time in seconds. This corresponds to a flight speed of $13.4 m/s$, or $48.2 km/h$.

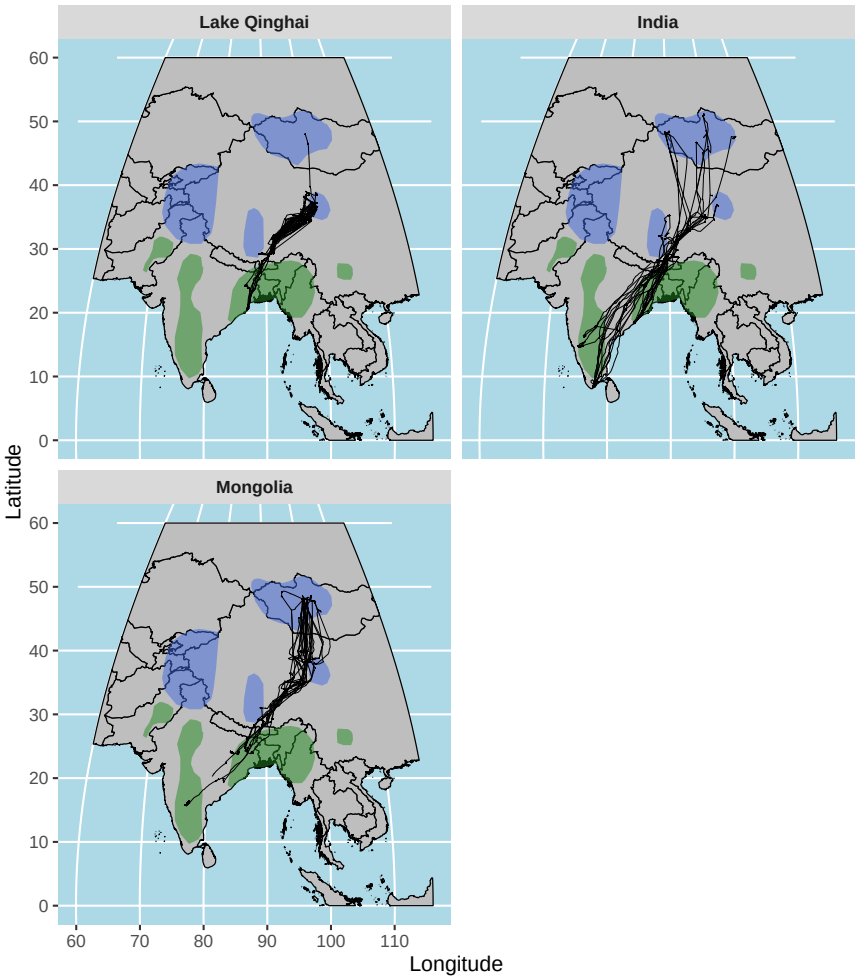

Figure S 1. Tracking data of bar-headed geese with individual GPS-trajectories of all individuals for the study period. The breeding range is shown in blue colour, the wintering range in green (BirdLife International and NatureServe, 2013).

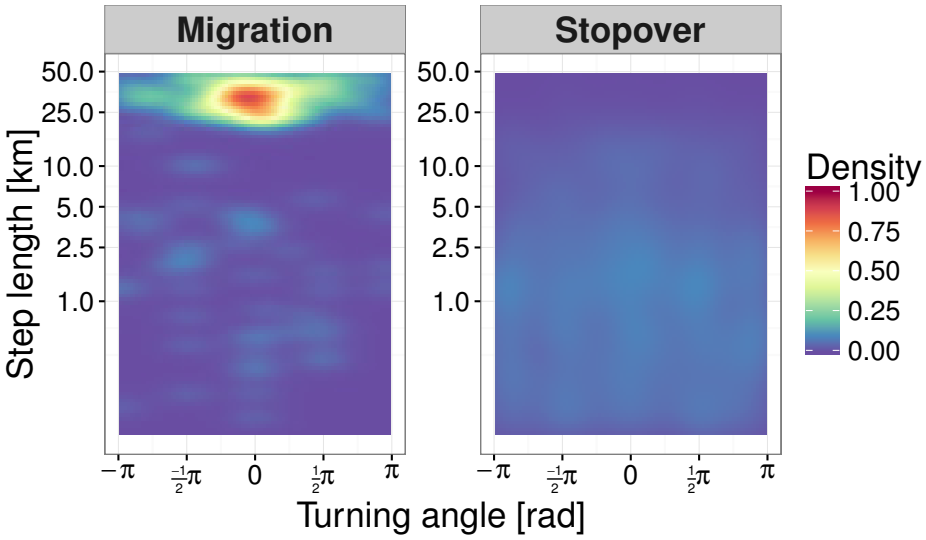

Figure S 2. This figure shows 2D-distributions of step length and turning angle for both movement modes, migration and stopover. The time lag between relocations and thus between estimates for speed and turning angle was two hours.

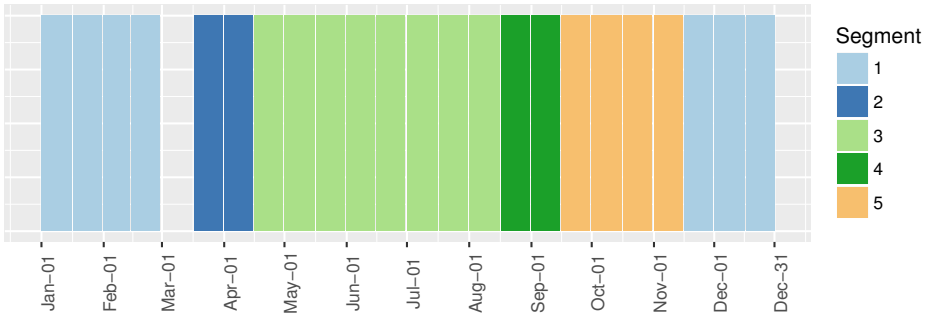

Figure S 3. Temporal segmentation by habitat use derived from the tracking data of all individuals. The colouring of segments is arbitrary and does not reflect similarities between segments.

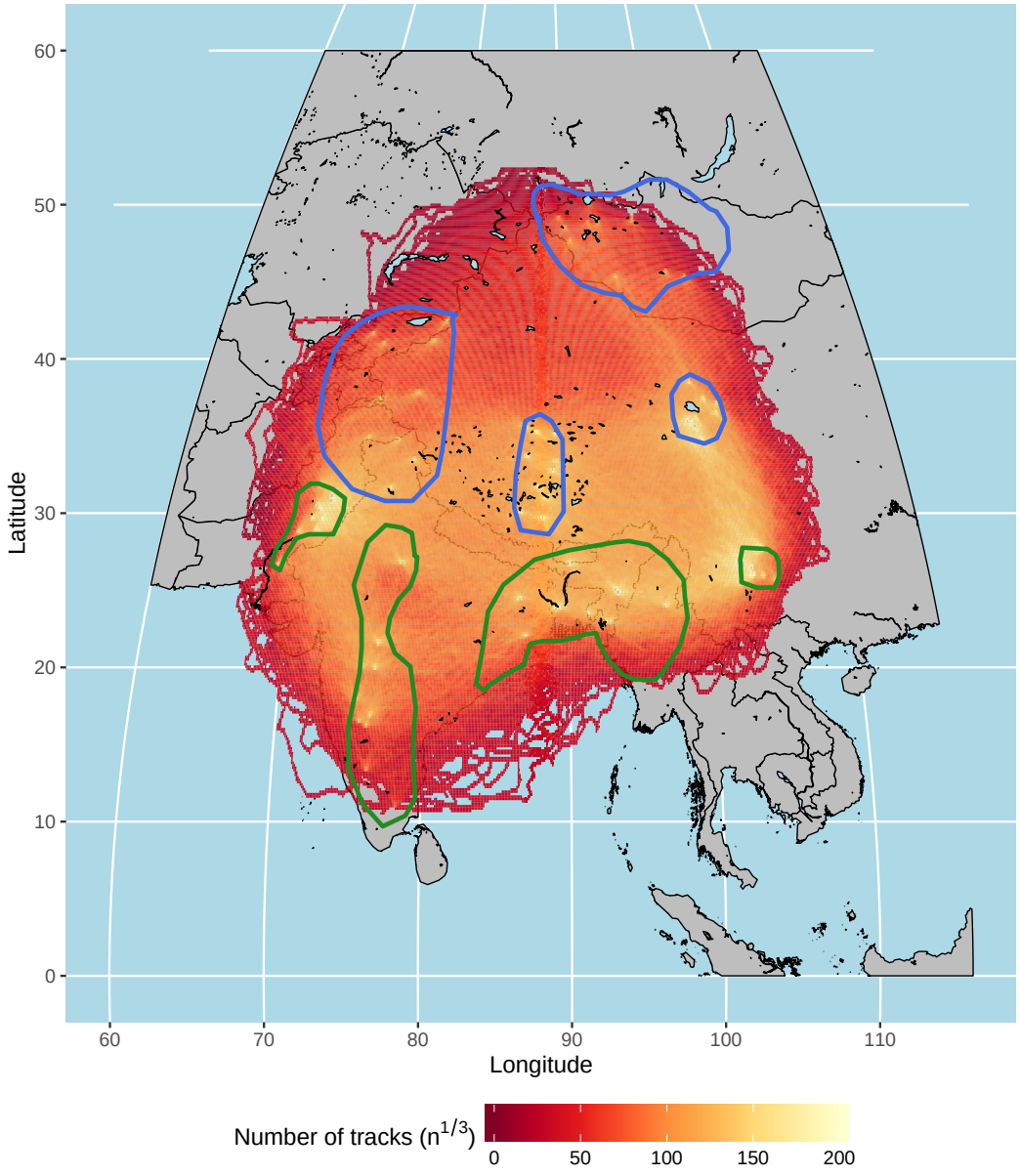

Figure S 4. This map displays the hitchhiker's guide to the migration of bar-headed geese caught at Lake Qinghai. Shown are only tracks with positive fitness, and the colouring illustrates the density of tracks. Blue polygons illustrate the native breeding range, green polygons the native wintering range of the species.

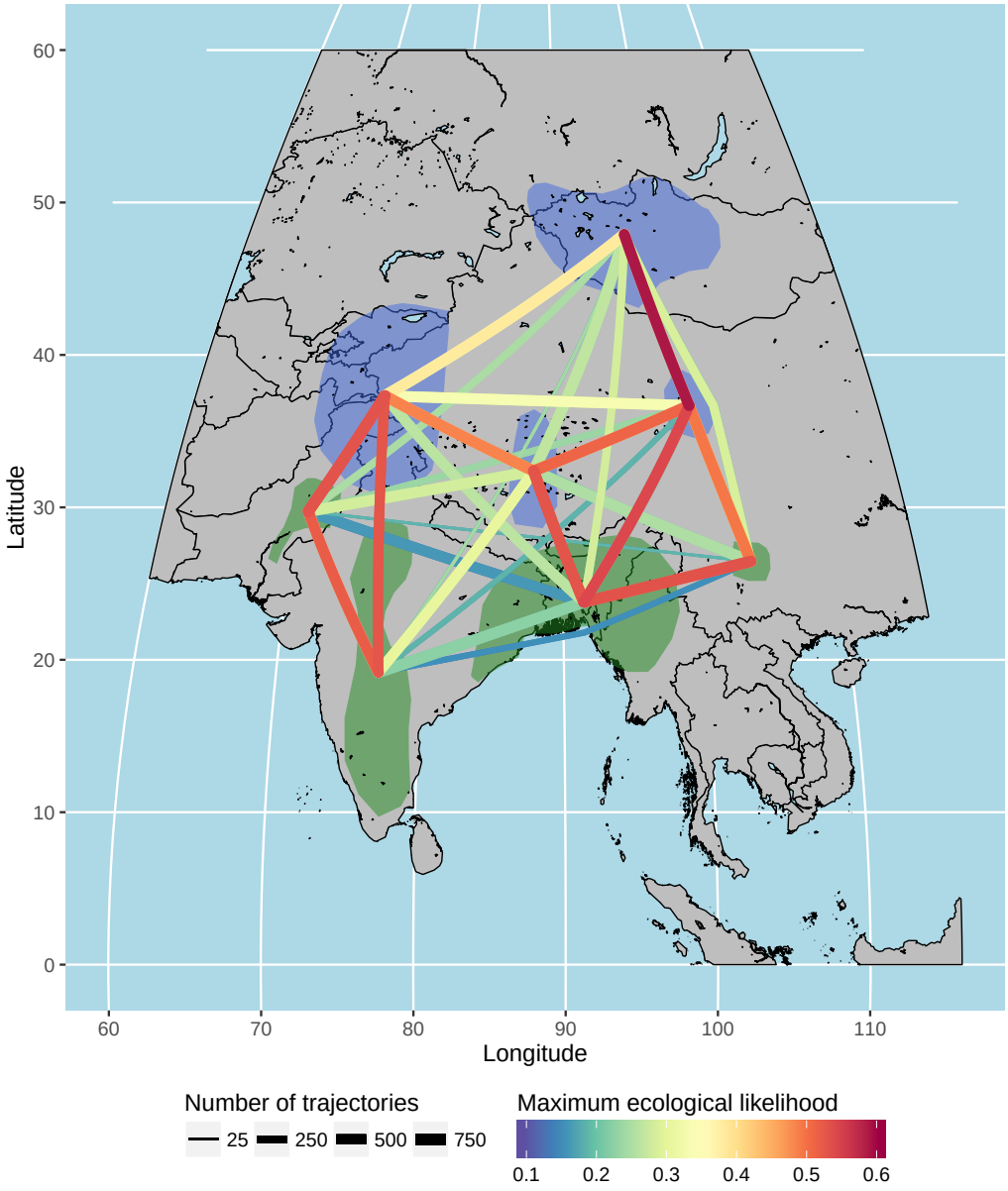

Figure S 5. Here we show the maximum connectivity between range fragments of bar-headed geese. We summarised the ecological likelihoods for all pairwise range fragment trajectories using the maximum ecological likelihood. The thickness of edges represents the sample size. Blue polygons show the native breeding area of the species, green polygons the native wintering range. Long edges are curved for sake of visibility.

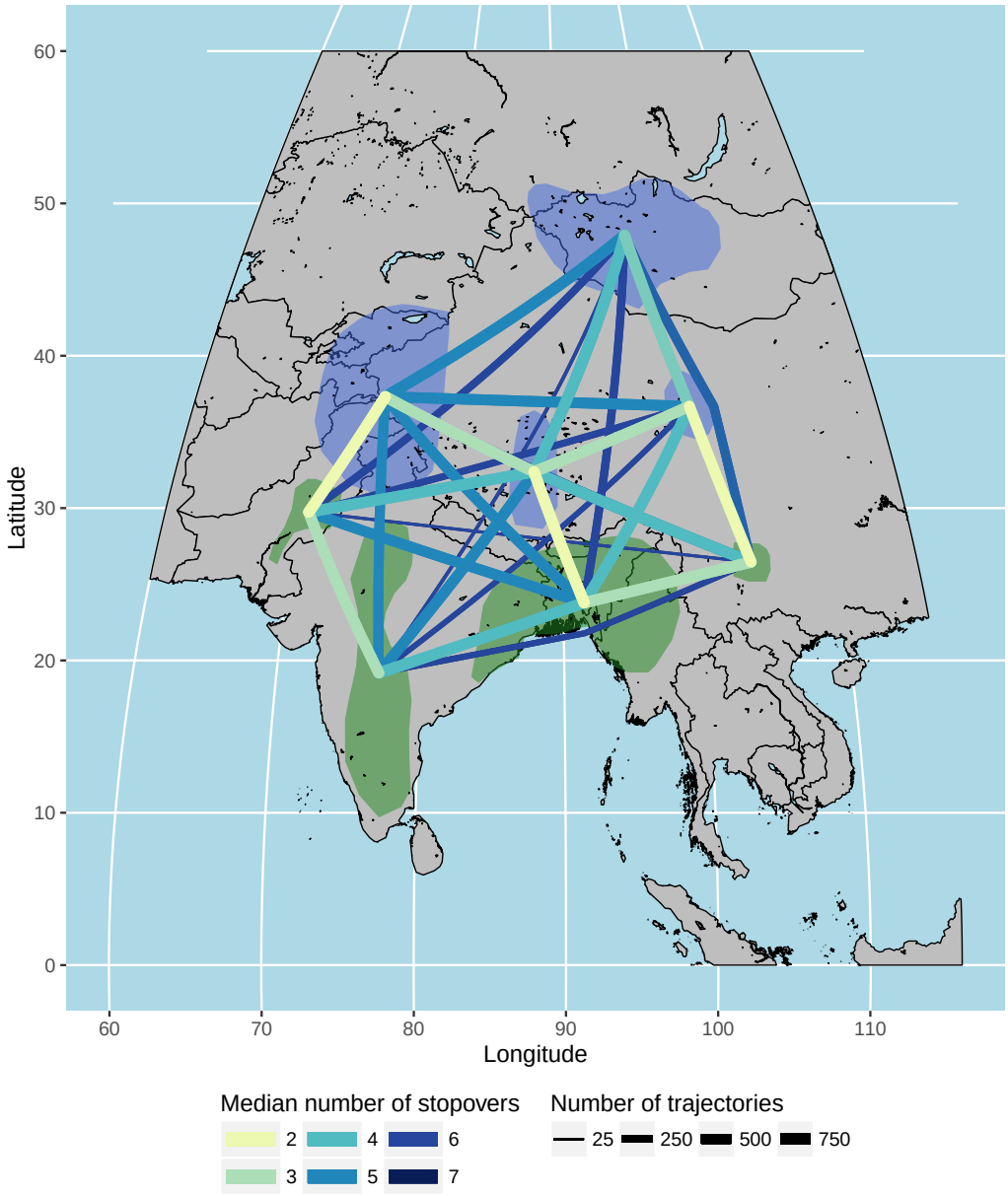

Figure S 6. Here we show the median number of stopovers that are predicted to occur between range fragments as derived from simulated migrations. Blue polygons show the native breeding area of the species, green polygons the native wintering range. Long edges are curved for sake of visibility.

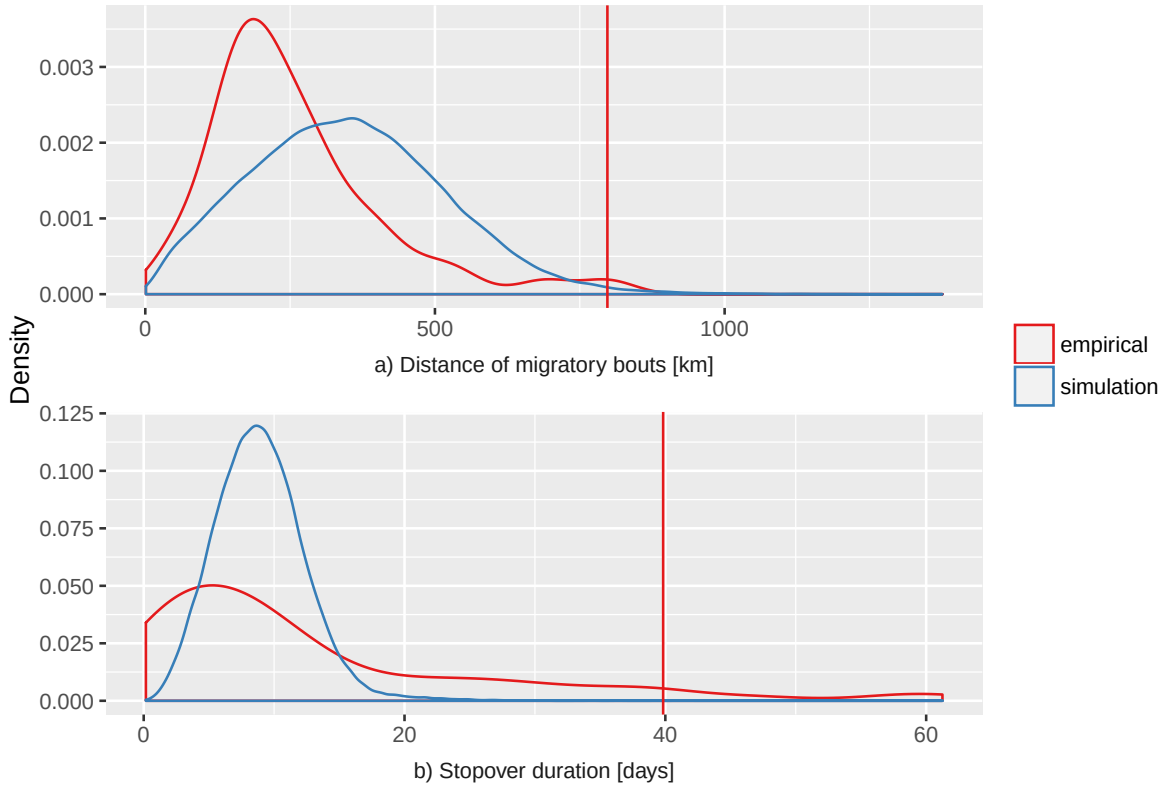

Figure S 7. Here we show the comparison between the observed and simulated properties of trajectories. The blue density curves show the cumulative distances flown between stopover sites (a, $\sum_{i=0}^{t} dm$) and the amount of time spent at stopover sites (b, Ts) for the simulated trajectories. The red density curves display the respective observations in the empirical trajectories, and the red vertical lines the values used for the maximum migratory distance, dm_{max}, and 95% of stopover duration, Ts_{max}.

References

Breiman, L (2001). Random Forests. *Mach Learn* 45:5–32.

Dodge, S, Bohrer, G, Weinzierl, R, *et al.* (2013). The Environmental-Data Automated Track Annotation (Env-DATA) System: Linking Animal Tracks with Environmental Data. *Movement Ecol* 1:3.

Garriga, J, Palmer, JRB, Oltra, A, *et al.* (2016). Expectation-Maximization Binary Clustering for Behavioural Annotation. *PLOS ONE* 11:e0151984.

International, B and NatureServe (2013). *Bird Species Distribution Maps of the World. Version 3.0.*

Kranstauber, B, Kays, R, LaPoint, SD, *et al.* (2012). A Dynamic Brownian Bridge Movement Model to Estimate Utilization Distributions for Heterogeneous Animal Movement. *J Anim Ecol* 81:738–746.

Lehner, B and Döll, P (2004). Development and Validation of a Global Database of Lakes, Reservoirs and Wetlands. *J Hydrol* 296:1–22.

Liu, C, Berry, PM, Dawson, TP, *et al.* (2005). Selecting Thresholds of Occurrence in the Prediction of Species Distributions. *Ecography* 28:385–393.

Olson, DM, Dinerstein, E, Wikramanayake, ED, *et al.* (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* 51:933.

Rousseeuw, PJ (1987). Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J Comput Appl Math* 20:53–65.

van Toor, ML, Newman, SH, Takekawa, JY, *et al.* (2016). Temporal Segmentation of Animal Trajectories Informed by Habitat Use. *Ecosphere* 7:e01498.

Wessel, P and Smith, WHF (1996). A Global Self-Consistent, Hierarchical, High-Resolution Shoreline Database. *J Geophys Res* 101:8741–8743.